Published online March 12, 2007

Assessing Microbial Community Diversity Using Amplicon Length Heterogeneity Polymerase Chain Reaction

DeEtta K. Mills*
Dep. of Biological Sci. and Intl. Forensic Research Inst.
Florida International Univ.
Miami, FL 33199

James A. Entry
USDA-ARS
Northwest Irrigation and Soils Research Lab.
Kimberly, ID 83341

Patrick M. Gillevet
Dep. of Environmental Science and Policy
George Mason Univ.
Manassas, VA 20110.

Kalai Mathee
Dep. of Biological Sci. and Intl. Forensic Research Inst.
Florida International Univ.
Miami, FL 33199

It is thought that a microbial community is an assemblage of organisms, genes, and gene functions. Transient, acute signals such as excessive nutrient loads or disturbance and chronic signals such as seasonal temperature or rainfall impact the total environmental system. The goal of many microbial ecologists is to determine if a finely resolved study of microbial dynamics can be used as a large-scale biosensor to follow diversity patterns in the environment. With the development of new genomic tools, community-level studies have been designed that can interrogate the finer details of the biological components of a given habitat. Amplicon length heterogeneity polymerase chain reaction (LH-PCR) interrogates the hypervariable domains of the ribosomal small-subunit genes and separates these domains based on the naturally occurring sequence lengths of DNA. The amplicons are phylogenetically relevant in that the various amplicons generated can be directly associated with specific taxonomic sequences archived in the databases. The application of the LH-PCR technique as a monitoring tool for microbial ecology has been shown to enhance and extend the current understanding of the dynamics of microbial communities in their specific environments.

Abbreviations: bp, base pair; DGGE (TGGE) denaturing (or temperature) gradient gel electrophoresis; IGS, intergenic spacer; ITS, internally transcribed spacer; LH-PCR, length heterogeneity-polymerase chain reaction; rRNA, ribosomal ribonucleic acid; T-RFLP, terminal restriction length polymorphism; V1, V2, V3, variable domain 1, variable domain 2, variable domain 3 of the 16S ribosomal molecule.

Soil microorganisms are of great scientific interest because of their universal presence and functioning within all terrestrial ecosystems. They are responsible for organic matter degradation and the mineralization of essential plant nutrients and are therefore a vital link in the global biotransformation of nutrients. Microorganisms exhibit an impressive diversity in their metabolic activities and in their interactions with other microbes, plants, and animals. They often respond more quickly to perturbations to their habitats and may be the most sensitive indicator of the impact of anthropogenic activities (Xing et al., 1997; Degens et al., 2001).

To date, only a small fraction of the world's microbes have been phylogenetically identified (Pace, 1997; Hugenholtz et al., 1998). In the past this was due to the inability to cultivate the vast majority in the laboratory. As a result, in the discussion of ecosystem processes, the microbes were often placed in a "black box," inputs and outputs measured, but most of the processes that go on inside "the box" were based on inference (Caroll and Wicklow, 1992; Brussaard et al., 1997; Wolters et al., 2000; Swift et al., 2004). With the application of molecular tools to microbiology, however, culture-independent methods have added greatly to this phylogenetic information. The debate over how many bacterial species exist in nature is still ongoing but some have estimated the range of bacterial species richness and diversity in a gram of soil to be from 6400 to 38 000 species (Curtis et al., 2002).

Information gained from studies involving laboratory cultures is extremely important in the understanding of microbial physiology but cannot always be extrapolated to the organism's role or response within its natural habitat. Microbes exist in nature as complex, community-oriented entities responsible for driving processes that define and shape their surrounding environments (Davey and O'Toole, 2000; Watnick and Kolter, 2000). Molecular tools now allow investigators to probe microbial community function and structure at increasingly finer resolution without cultivation.

Molecular Community Profiling

There are two general classes of DNA polymorphisms that can be used to differentiate organisms (or communities) from one another: (i) base sequence polymorphisms within a DNA strand and (ii) insertions and deletions of bases that lead to

Soil Sci. Soc. Am. J. 71:572–578
doi:10.2136/sssaj2006.0147
Received 7 Apr. 2006.
*Corresponding author (millsd@fiu.edu).

length polymorphisms within a DNA strand. Because microbial diversity is so vast, several molecular approaches are often used by investigators to decipher the diversity and the dynamics of a microbial community. Ubiquitous molecular markers are commonly chosen to survey whole communities based on a conserved gene sequence. For microbial communities, the marker of choice is the highly conserved ribosomal genes (Embley and Stackebrandt, 1996) or other sites such as the intergenic spacer (IGS) or internally transcribed spacer (ITS) regions between or within ribosomal operons (Bourque et al., 1995; Chun et al., 1999). The gold standard for assessing sequence polymorphisms and phylogenetic diversity is the construction of clone libraries and high-throughput sequencing (Venter et al., 2004). These methods are most often used to discover which species are present within a particular environment.

Inferences about environmental impacts or manipulations can be made, however, using molecular tools that rapidly assess or profile the *dynamics* of a community without sequencing large clone libraries. These methods by themselves cannot detect which organisms may be responding to a disturbance or treatment but only that there was a change in the community structure. Profiling methods are designed to show an effect on a community or differences between communities but do not provide *direct* phylogenetic information.

The individual amplicon lengths can often be assigned phylogenetic status when combined with sequenced clone libraries. For example, using clone libraries, Suzuki et al. (1998) verified that LH-PCR amplicons ranging from 312 to 328 base pairs (bp) in a picoplankton community were most often associated with the Alphaproteobacteria group. Similarly, Sekar et al. (2006) used LH-PCR profiling and clone libraries to identify taxa representative of particular amplicons in their LH-PCR profiles. By virtually aligning the primer sequences to the cloned sequences, amplicons from the profiles could be directly associated with particular organisms found in the clone library. Others have used virtual alignments of archived sequences to calculate amplicon lengths and associate those lengths with certain taxa (Crosby and Criddle, 2003; Kent et al., 2003; Matsumoto et al., 2005). For any profiling technique used, however, a particular amplicon length can be associated with more than one taxon; therefore, caution must be used when trying to infer the true taxonomic affiliation of an amplicon based on a bioinformatics approach without verification by direct sequencing of sample-specific clone libraries.

Profiling methods that target base changes in DNA sequences are terminal restriction fragment length polymorphisms (T-RFLP) (Moeseneder et al., 1999; Simpson et al., 2002) and denaturing (or temperature) gradient gel electrophoresis (DGGE/TGGE) (Muyzer et al., 1993; Torsvik et al., 1996; Felske et al., 1998). Techniques that use the naturally occurring sequence length-based differences to distinguish between communities include: amplicon length heterogeneity PCR (LH-PCR) (Suzuki et al., 1998); automated rRNA intergenic spacer analysis (Ranjard et al., 2001); and the IGS-PCR (Fisher and Triplett, 1999) or ITS-PCR (Lord et al., 2002). The focus of this mini-review is on the length heterogeneity profiling and its application as a monitoring tool in various matrices such as soil, sediment and water.

Amplicon Length Heterogeneity Polymerase Chain Reaction: The Technique

Microbes can be distinguished based on the natural length polymorphisms that occur due to insertions and deletions of bases within genes or gene operons. Amplicon LH-PCR interrogates hypervariable domains, most frequently the ribosomal small subunit (*rrn*), and produces a DNA fingerprint or profile. This fluorescent-based PCR method is straightforward and, with optimization, highly reproducible (Suzuki et al., 1998; Mills et al., 2003).

The first step in LH-PCR is to extract the community DNA from the sample. The DNA extraction methods can vary, but one that has been proven to uniformly disrupt and produce PCR-ready DNA from most sample types is Q-BioGene's Fast DNA Spin Kit for Soil (Mills et al., 2003). The extracted metagenomic DNA is quantified by fluorimetry since different starting concentrations of DNA in a PCR reaction can bias interpretation of the data (Ritchie et al., 2000; Mills et al., 2003). Universal primer pairs are most often used with the forward primer having been labeled with a fluorescent dye. Products are separated by size on a genetic analyzer and the fluorescence captured by the instrument's software. The fluorescence data are converted into chromatographic profiles called electropherograms. A fluorescent internal size standard is run with each sample and is used to size the amplicon lengths in base pairs. The intensity (height) or area under the peak in the electropherogram is proportional to the relative abundance of that particular amplicon. Each matrix type should be optimized to establish the correct concentration of metagenomic DNA, the choice of universal (or species-specific) primer sets, the fluorescent tag on the primers and the ratio of PCR reagents to use. In addition, the optimal number of PCR cycles needs to be established to minimize template reannealing and other PCR artifacts (Suzuki and Giovannoni, 1996) that could confound data interpretation.

Technical Issues, Advantages, and Limitations of the Technique

Each technique has its advantages and limitations and these factors need to be considered when choosing which method is best suited for a particular study (Hermans et al., 1995; Wintzingerode et al., 1997; Keim et al., 2000; Alves et al., 2002; Crosby and Criddle, 2003; Mills et al., 2003). For example, there is an inherent bias to *any* molecular technique that uses the 16S rRNA gene operons for community analyses. Prokaryotes have variable numbers of copies of the *rrn* operon that can range from one copy for organisms such as *Chlamydia trachomatis* and *Bradyrhizobium japonicum* to as many as 13 for *Bacillus cereus* or from 10 to 15 copies for various *Clostridium* spp. (http://rrndb.cme.msu.edu [verified 21 Dec. 2006]). When using these methods for whole-community analysis, only the minimal richness or differences between profiles are measurable due to variation in operon copy numbers (Farrelly et al., 1995; Klappenbach et al., 2000) and co-migrating amplicons from phylogenetically distinct taxa (Suzuki et al., 1998).

Given the technical limitations, LH-PCR can still provide insight into the community organization without the costly and labor-intensive construction of clone libraries and DNA

sequencing analysis. The advantages of LH-PCR are: (i) the PCR products do not require post-PCR clean up or restriction enzyme digestion and can be directly loaded onto the genetic analyzer; (ii) replicate profiles are highly reproducible if the DNA extraction method is robust and metagenomic DNA has been quantified; and (iii) it can quickly evaluate community changes within or between treatments or over time and space. Limitations are: (i) the profiles often have contiguous amplicon distributions that are sometimes difficult to resolve with the current software; (ii) one amplicon can represent more than one taxon that are phylogenetically distinct but produce the same length amplicon; and (iii) individual fractions or peaks cannot be collected or separated, which means whole-community clone libraries need to be constructed. As with any profiling technique, each investigator must identify and quantify sources of methodological and systematic errors when using LH-PCR to study microbial community dynamics.

Length Heterogeneity and the Ribosomal Operon

One of the first published investigations using LH-PCR was in Suzuki et al. (1998). This study used LH-PCR to analyze the relative 16S rDNA amplicon frequencies from picoplankton found in marine waters off the coast of Oregon. The investigators used two sets of universal primers to analyze the first two or three hypervariable domains at the 5′ end of the 16S rRNA gene and also amplified and cloned the full-length 16S rRNA community genes. They tested for PCR bias by using LH-PCR and varying the number of PCR cycles. By testing and limiting the number of PCR cycles used in the amplification process, any kinetic bias associated with PCR could be reduced and the results proved to be highly reproducible. When they reamplified the DNA that had been used for the clone library construction, they observed an obvious kinetic bias that was caused by template reannealing when performing 35 cycles vs. the optimal 25 cycles. Although many of the same peaks were present in the second LH-PCR profile, the relative contributions of the peaks had changed due to template reannealing with more cycles. This is an important point since community analyses using profiling data are often based not only on the presence or absence of an amplicon but on the relative contribution of that amplicon to the whole-community profile. They also compared the profiles from the clone libraries to the original LH-PCR profile and found them to be quite different. They concluded that this bias was probably due to the random selection of clones from the library and other inherent biases to PCR and sequencing. The results of this study were twofold: (i) it demonstrated a kinetic bias in template reannealing when amplifying complex natural communities, especially if too many PCR cycles were performed; and (ii) it confirmed that LH-PCR was a valid, reproducible method for quickly assessing the overall complexities of a microbial community. The LH-PCR amplicon lengths were validated using the clone library sequences derived from the same sample. The relative proportion and amplitude of peaks corresponded to the percentage of identified clones in the libraries.

Ritchie et al. (2000) performed one of the first published applications of LH-PCR in a soil matrix. In this study, they surveyed four different soils under different land management practices and performed both LH-PCR and fatty acid methyl ester (FAME) analyses. The FAME technique uses cellular lipids to profile viable microbial communities and has been extensively used in many studies (Vestal and White, 1989; Cavigelli et al., 1995; Hanson et al., 1999; Stephen et al., 1999; Glucksman et al., 2000; Hutchinson et al., 2000). Extracted soil DNA was used to generate LH-PCR profiles from the first two hypervariable domains of the 16S rRNA genes. Reproducibility of the technique was studied at the plot, DNA extraction, amplification, and analysis steps. When the coefficients of variation were compared for the averaged data, the greatest variability was seen at the plot level (14%), which was expected due to sampling and the natural heterogeneity of soil within a site. Based on principle component analyses of the averaged peak data and FAMEs, Ritchie et al. (2000) demonstrated that both LH-PCR and FAME were capable of distinguishing the soil microbial communities associated with the four soil types. They concluded, however, that the LH-PCR was better at discriminating subtle differences between tillage practices. The clone libraries that were sequenced, however, produced conflicting phylogenetic identities when they attempted to correlate LH-PCR and FAME data. While FAME data showed a negative correlation to Gram-positive microbes, the sequence data associated with a particular amplicon identified the clone as a Gram-positive organism. This conclusion was further weakened, however, by the fact that they cloned only seven of 33 bacterial clones. They used LH-PCR to screen the clone libraries that allowed picking a particular amplicon length to sequence. By doing this preselection of clones and sequencing so few, they biased their conclusions assuming that one amplicon length represented only one organism, while in fact, it has been shown not to be the case (Mills et al., 2003, Sekar et al., 2006).

Our group has used LH-PCR to test its ability to distinguish between microbial community patterns from contaminated soils (Mills et al., 2003) and from the same soil type under different land management practices (Mills et al., 2006). In the first study (2003), LH-PCR was compared with T-RFLP to assess which method was better able to monitor the impact nutrient amendments may have on soils contaminated with hydrocarbons. Bench-scale bioreactors were amended with optimal inorganic nutrients (treatment) or left unamended (controls). Contaminated soils were slurried and placed in the air-lift bioreactors. Halfway through the 30-d monitoring period, the bioreactors were spiked with Arabian light crude oil. The same DNA extracts were used to test both profiling methods. The LH-PCR technique used the same universal primer pair as Suzuki et al. (1998) and Ritchie et al. (2000) to amplify the first two hypervariable domains, V1 + V2, of the 16S rRNA genes. Another universal primer pair was used to amplify the entire 16S rRNA genes for the T-RFLP and both the forward and reverse primers were labeled. All other experimental conditions remained uniform across experiments. The LH-PCR technique was better able to follow the dynamics in the bioreactors for several different reasons. First, it was found that LH-PCR was highly reproducible and was not plagued with post-PCR manipulation biases. Replicate enzymatic digestions (even with different restriction enzymes) often led to partial digests that produced a T-RFLP profile that was not reproducible. Blocking of restriction sites by inhibi-

tors or mixed templates may have contributed to this technical problem. Since LH-PCR did not involve post-PCR digestions, these technical artifacts were not an issue. Also, fewer diagnostic terminal fragment lengths were produced by T-RFLP than in the LH-PCR in this study. Coincidently, many of the organisms produced the same length terminal restriction fragments and thus produced a less complex profile. Clone libraries were constructed to verify some of the bacteria associated with both LH-PCR and T-RFLP peaks and the libraries were dominated by known hydrocarbon degraders (e.g., Alpha- and Gammaproteobacteria). All clones were aligned with the respective primer pairs used for both T-RFLP and LH-PCR and the virtual amplicon lengths were obtained. Many of the terminal fragments produced by the restriction digests were not able to be resolved due to higher signal to noise in the T-RFLP output or the inability to distinguish the shortest terminal fragments because they were lost in the primer peak background noise. It was concluded that, for this particular study, LH-PCR was better able to monitor the dynamics within the bioreactors than was T-RFLP since more unique amplicons were produced based on length heterogeneity than restriction site sequence heterogeneity. This study demonstrates the need for each investigator to carefully choose the technique that best provides data that will adequately reflect the behavior of the community dynamics.

Other studies have used LH-PCR to follow the dynamics of microbial communities in various habitats. Carlson et al. (2002) used LH-PCR to monitor the responses of bacterioplankton communities to nutrient amendments. Significant temporal changes were seen in the bacterial community profiles when labile dissolved organic carbon (DOC) and inorganic N and P were added to seawater cultures. The labile DOC also enhanced the utilization of seasonal semilabile DOC. Since the overall change in biomass and cell count abundances remained relatively stable, they concluded that there was a redistribution of population sizes among the dominant and rare members when profiles were compared with time zero (Carlson et al., 2002).

Bernhard and Field (2000) adapted LH-PCR to track nonpoint-source fecal pollution in Oregon coastal waters. They amplified human and bovine fecal samples with species-specific primers developed for host-specific fecal anaerobes, *Bacteroides–Prevotella* spp. and *Bifidobacterium* spp. They identified a peak at 276 bp that was bovine specific with no amplification of human host *Bacteroides–Prevotella* DNA. This led them to conclude that the nonpoint-source pollution was from farming practices and not leaking septic tanks. They demonstrated that host-specific LH-PCR was a rapid and sensitive method that had great potential for tracking nonpoint-source pollution, pathogens, and genetically engineered and released bacteria in the environment (Bernhard and Field, 2000), thus expanding the use of LH-PCR for environmental monitoring. This same approach could be applied to soil. By choosing species-specific primers, it would be possible for LH-PCR to track pathogens based on their associated amplicon lengths in manure-amended soils. This would allow rapid assessment of pretreatment strategies (e.g., composting) of manure handling by monitoring the survival or elimination of pathogens from the source before application to agricultural fields.

Tiirola et al. (2003) used LH-PCR to track the community dynamics in an aerobic suspended carrier biofilm reactor set up to treat whitewater waste from a pulp and paper mill (Tiirola et al., 2003). The profiles indicated that *Bacillus* spp. were the dominant populations detected after alkaline shock to the community. When allowed to recover, however, profiles similar to the original microbial community were obtained and monitored the community's recovery after shock treatment. Other studies have used LH-PCR to profile communities in dairy whey starters (Lazzi et al., 2004), in enrichment cultures from sheep ruminal fluid for the consortium's ability to degrade pyrrolizidine alkaloids (Lodge-Ivey et al., 2005), and in assessing agricultural tillage practices on microbial populations (Mills et al., 2003). All of these applications demonstrate the usefulness of LH-PCR profiling or any profiling method as a rapid monitoring tool of community changes.

Application of the Technique to Other Genes

Length heterogeneity is not restricted to only structural genes such as the ribosomal operons but, in many cases, noncoding or transcribed intergenic regions within operons can produce sequence length polymorphisms. For example, autotrophic NH_4–oxidizing bacteria transform NH_3 to hydroxylamine. The essential enzyme for that transformation, ammonia monooxygenase, is encoded in a three-gene operon, the *amo* operon. In a study by Norton et al. (2002), the intergenic region between the *amoC* and the *amoA* genes was found to have length heterogeneity and could potentially be used to profile NH_3–oxidizing bacteria from environmental samples. The amplicons ranged in size from 277 bp for *Nitrosomonas europaea* 19178 to 553 bp for *Nitrosospira* sp. 39019. In preliminary studies using the same published primers, our group profiled the NH_3–oxidizing community from composted and fresh dairy manure (Fig. 1). The raw data (electropherogram) showed similar lengths to the data of Norton et al. (2002) within the *amo* intergenic region. No *amo* signal was found in the control (no amendment) samples under the conditions of this run. While optimization of the *amoC–amoA* LH-PCR is ongoing in our laboratory, it appears that this operon has good potential to be used as a LH-PCR marker for this group of specialized organisms. In an ongoing mesocosm study involving different fertilizer regimes, preliminary data have shown the *amo* signal to increase from below detection at time zero to 10- to 20-fold at later time points in fertilized (NH_4NO_3 addition) mesocosms (data not shown). In addition, we are exploring other genes and operons that would be potential candidates for this molecular technology. This would allow profiling of functional gene populations within a defined structural community. The relative ease, reproducibility, and robustness of the LH-PCR method increases its application to further exploit the natural sequence length variations within genes and operons and among organisms. This makes LH-PCR a useful molecular tool in the study of whole-community dynamics in multiple model systems and environments.

Analysis of Length Heterogeneity–Polymerase Chain Reaction Profiles

When microbial community profiling studies first began to be published, much of the data were presented as raw electropherograms or denaturing gels (Suzuki et al., 1998; Moeseneder et

al., 1999; Ritchie et al., 2000) showing the presence or absences of amplicons. Investigators could profile the microbial communities using different methods but were still perplexed about the best method to use for analyzing the data (Watve and Gangel, 1996; Hughes et al., 2001; Hill et al., 2003).

As molecular microbial ecology has advanced, so has the need to find more suitable metrics with which to analyze the data. In two recent studies by our group, LH-PCR was used to query which hypervariable domain or combination of 16S rRNA gene domains was the best molecular marker (Mills et al., 2006; Yang et al., 2006). In the first study, data from Idaho natural sagebrush and irrigated moldboard plowed sites were used to compare univariate and multivariate analyses. Using standard ecological indices (i.e., richness, diversity, and evenness), the combinations of three hypervariable domains (V1, V3, and V1 + V2) or the combined V1 and V3 domains, were found to be the most discriminatory. Since most profile data are not normally distributed, a comparison was made using multivariate Bray–Curtis similarity and multidimensional scaling (MDS) and these were found to be better metrics to ordinate and cluster the LH-PCR community profiling data. The depth and disturbance patterns of the communities using MDS was strongly reflected in the separation of data points on the MDS plot, whereas the diversity measures collapsed that data to a one-dimensional index (Mills et al., 2006). Bernhard et al. (2005) also used LH-PCR profiles and MDS to map the distribution of microbial communities that spanned fresh and brackish waters in northwestern estuaries and tributaries. This multivariate approach appears to be better suited to analyze complex data that are not normally distributed.

Using supervised learning tools such as K-nearest neighbor methods (KNN) and support vector machines (SVM), Yang et al. (2006) tested the ability of these computational tools to distinguish sediment and soil sample communities based on LH-PCR profile data. Four hypervariable domains (V1, V3, V1 + V2, and V9) of the 16S rRNA genes were tested for their ability to classify the samples by treatment, depth of sampling, season, or location. Supervised learning tools train on a set of known samples and learn to classify unknown samples based on labeled feature vectors obtained from the data when added to the trained set. Chesapeake Bay sediment samples and Idaho agricultural soils were used to create LH-PCR profiles using various domains within the 16S rRNA genes. Both of the classifiers (KNN and SVM) were able to accurately classify the Idaho soils into their respective treatment and depth categories and predict which soil group new data belonged to when added to the trained set. It also found that by combining the data from V1, V1 + V2, and V3, the LH-PCR profile data complemented each other and gave 100% accuracy in the classification of the samples. When the V9 hypervariable domain data were added to the training set, it provided no additional value and, in fact, actually lowered the ability of the SVM to accuracy classify the soil. When the algorithm was run independently on the Chesapeake Bay samples, the SVM classifier was better able to classify the sediments than was the KNN. The averaged accuracy was only 83% based on sample location, but SVMs were able to distinguish seasonal differences with an accuracy range of 92 to 95%. The less accurate results with the Chesapeake Bay community profile data suggested greater similarity within the bay microbial communities than was seen in the agricultural managed soils.

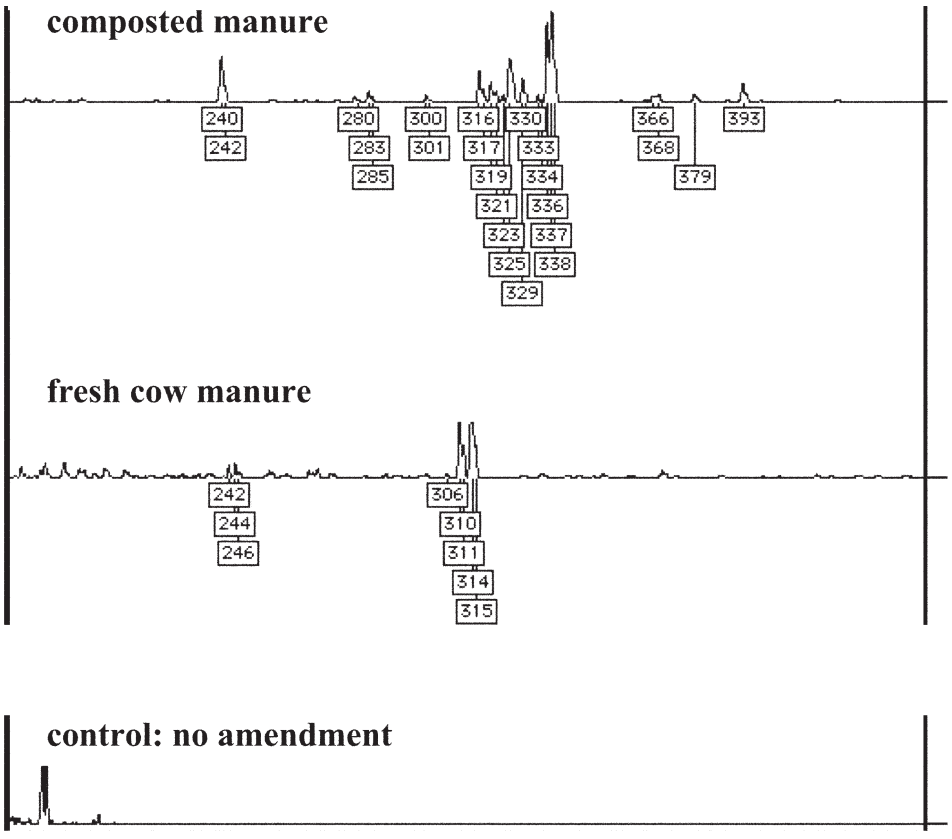

Fig. 1. Length heterogeneity–polymerase chain reaction applied to the intergenic region between the *amo*C and the *amo*A genes of the ammonia monooxygenase operon. The top panel is the electropherogram from soil treated with composted manure, the middle panel is the electropherogram from fresh cow manure, and the bottom panel is unamended soil (no manure) where the genes are either below detection or absent. The numbers in the box are the lengths of the amplicons in base pairs.

The use (or misuse) of traditional ecological indices such as diversity and evenness to study microbial community profiles will remain a major point to consider when performing microbial ecology studies. Perhaps other metrics or computational algorithms need to be developed and tested that will give more insight into the community's genetic structure and function.

CONCLUSIONS

Since a microbial community is an assemblage of organisms, genes,

and gene functions, transient, acute signals such as excessive nutrient loads or chronic signals like seasonal temperature or rainfall impact the entire assemblage. Amplicon LH-PCR is one of several DNA profiling techniques that can be applied to the study of microbial community dynamics and diversity. As with any profiling technique, it has its advantages and its limitations. Each model system or ecosystem matrix needs to be assessed for the appropriate use of this technique in answering the hypotheses set forth. In combination with clone libraries and correlations with ecosystem drivers, LH-PCR can provide a rapid assessment of the effect that anthropogenic and environmental stressors or natural fluxes may have on the microbial populations.

In concert with our ability to study microbial communities in situ is the need to refine our analysis of community data. Complex data sets that do not fit normal distributions need to be analyzed using nonparametric measures vs. more traditional univariate diversity indices (Bernhard et al., 2005; Mills et al., 2006). New algorithms or application of bioinformatics tools need to be tested for the ability to better understand the encompassing complexities of the metagenome that exists at the microscopic scale within different environmental landscapes (Yang et al., 2006).

Complex ecological systems emerge from interactions between organisms and physiochemical phenomena operating across broad scales of time and space. The vast majority of the planet's net primary productivity flows through microorganisms, fueling key steps in the biogeochemical cycles of C, N, S and important trace metals, as well as most other elements. Because microbial populations operate at spatiotemporal scales far removed from typical human perception, it has been difficult to investigate their role in complex systems. Molecular techniques that can help increase our understanding of the genotypic diversity of microorganisms, their contribution to specific activities, and their distribution in their environments are now available.

To fully understand the overwhelming hidden diversity that is present within microbial communities and their associated ecosystems, we will need concerted interdisciplinary efforts, greater dissemination of information, and perhaps even higher resolution methods to understand these complexities. Studies that encompass both the biotic and abiotic interactions and responses will be important. Future studies that analyze function vs. structure or that can monitor several trophic levels at one time will lead to even a better understanding of the mutualistic, synergistic, and antagonistic interactions within a system. Together with traditional microbiological approaches, sequencing of clone libraries, and new computational tools, profiling methods such as LH-PCR, T-RFLP, or DGGE will continue to help narrow the knowledge gap in our understanding of the microbial community's role within the global ecosystem.

ACKNOWLEDGMENTS

This review was funded in part by an NSF ADVANCE Fellowship (Award no. 0340695) to D.K. Mills and in part by USDA Agricultural Research Service, Northwest Irrigation and Soils Research Lab., Kimberly, ID.

REFERENCES

Alves, A., O. Santos, I. Henriques, and A. Correia. 2002. Evaluation of methods for molecular typing and identification of members of the genus *Brevibacterium* and other related species. FEMS Microbiol. Lett. 213:205–211.

Bernhard, A.E., and K.G. Field. 2000. Identification of nonpoint sources of fecal pollution in coastal waters by using host specific 16S ribosomal DNA genetic markers from fecal anaerobes. Appl. Environ. Microbiol. 66:1587–1594.

Bernhard, A.E., D. Colbert, J. McManus, and K.G. Field. 2005. Microbial community dynamics based on 16S rRNA gene profiles in a Pacific Northwest estuary and its tributaries. FEMS Microbiol. Ecol. 52:115–128.

Bourque, S.N., J.R. Valero, M.C. Lavoie, and R.C. Levesque. 1995. Comparative analysis of the 16S to 23S ribosomal intergenic spacer sequences of *Bacillus thuringiensis* strains and subspecies and of closely related species. Appl. Environ. Microbiol. 61:1623–1625.

Brussaard, L., V.M.B. Pelletier, D.E. Bignell, V.K. Brown, W. Didden, P. Folgarait et al. 1997. Biodiversity and ecosystem functioning in soil. Ambio 26:563–570.

Carlson, C.A., S.J. Giovannoni, D.A. Hansell, S.J. Goldberg, R. Parsons, M.P. Otero, K. Vergin, and B.R. Wheeler. 2002. Effect of nutrient amendments on bacterioplankton production, community structure, and DOC utilization in the northwestern Sargasso Sea. Aquat. Microb. Ecol. 30:19–36.

Caroll, G.C., and D.T. Wicklow (ed.). 1992. The fungal community: Its organization and role in the ecosystem. Marcel Dekker, New York.

Cavigelli, M.A., G.P. Robertson, and M.J. Klug. 1995. Fatty acid methyl ester (FAME) profiles as measures of soil microbial community structure. Plant Soil 170:99–113.

Chun, J., A. Huq, and R.R. Colwell. 1999. Analysis of 16S–23S rRNA intergenic spacer regions of *Vibrio cholerae* and *Vibrio mimicus*. Appl. Environ. Microbiol. 65:2202–2208.

Crosby, L.D., and C.S. Criddle. 2003. Understanding bias in microbial community analysis techniques due to *rrn* operon copy number heterogeneity. Biotechniques 34:790–802.

Curtis, T.P., W.T. Sloan, and J.W. Scannell. 2002. Estimating prokaryotic diversity and its limits. Proc. Natl. Acad. Sci. 99:10494–10499.

Davey, M.E., and G.A. O'Toole. 2000. Microbial biofilms: From ecology to molecular genetics. Microbiol. Mol. Biol. Rev. 64:847–867.

Degens, B.P., L.A. Schipper, G.P. Sparling, and L.C. Duncan. 2001. Is the microbial community in a soil with reduced catabolic diversity less resistant to stress or disturbance? Soil Biol. Biochem. 33:1143–1153.

Embley, T.M., and E. Stackebrandt. 1996. The use of 16S ribosomal RNA sequences in microbial ecology. p. 39–62. *In* R.W. Pickup and J.R. Saunders (ed.) Molecular approaches to environmental microbiology. Ellis Horwood, London.

Farrelly, V., F.A. Rainey, and E. Stackebrandt. 1995. Effect of genome size and *rrn* gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. Appl. Environ. Microbiol. 61:2798–2801.

Felske, A., A.D.L. Akkermans, and W.M. deVos. 1998. Quantification of 16S rRNAs in complex bacterial communities by multiple competitive reverse transcription-PCR in temperature gradient gel electrophoresis fingerprints. Appl. Environ. Microbiol. 64:4581–4587.

Fisher, M.M., and E.W. Triplett. 1999. Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its applications to freshwater bacterial communities. Appl. Environ. Micorbiol.65:4630–4636.

Glucksman, A.M., H.D. Skipper, R.L. Brigmon, and J.W.S. Domingo. 2000. Use of the MIDI FAME technique to characterize groundwater communities. J. Appl. Microbiol. 88:711–719.

Hanson, J.R., J.L. Macalady, D. Harris, and K.M. Scow. 1999. Linking toluene degradation with specific microbial populations in soil. Appl. Environ. Microbiol. 65:5403–5408.

Hermans, P., M. Sluijter, T. Hoogenboezem, H. Heersma, A. van Belkum, and R. de Groot. 1995. Comparative study of five different DNA fingerprint techniques for molecular typing of *Streptococcus pneumoniae* strains. J. Clin. Microbiol. 33:1606–1612.

Hill, T.C.J., K.A. Walsh, J.A. Harris, and F.M. Bruce. 2003. Using ecological diversity measures with bacterial communities. FEMS Microbiol. Ecol. 43:1–11.

Hugenholtz, P., B.M. Goebel, and N.R. Pace. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol. 180:4765–4774.

Hughes, J.B., J.J. Hellmann, T.H. Ricketts, and B.J.M. Bohannon. 2001. Counting the uncountable: Statistical approaches to estimating microbial diversity. Appl. Environ. Microbiol. 67:4399–4406.

Hutchinson, G., K. Herrity, H. Malnick, and H.N. Shah. 2000. Long-chain cellular fatty acids; the database and its current applications in microbial

identification. Anaerobe 6:115–116.
Keim, P., L.B. Price, A.M. Klevytska, K.L. Smith, J.M. Schupp, R. Okinaka, P.J. Jackson, and M.E. Hugh-Jones. 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J. Bacteriol. 182:2928–2936.
Klappenbach, J.A., J.M. Dunbar, and T.M. Schmidt. 2000. rRNA operon copy number reflects ecological strategies of bacteria. Appl. Environ. Microbiol. 66:1328–1333.
Kent, A.D., D.J. Smith, B.J. Benson, and E.W. Triplett. 2003. Web-based phylogenetic assignment tool for analysis of terminal restriction fragment length polymorphism profiles of microbial communities. Appl. Environ. Microbiol. 69:6768–6776.
Lazzi, C., L. Rossetti, M. Zago, E. Neviani, and G. Giraffa. 2004. Evaluation of bacterial communities belonging to natural whey starters for Grana Padano cheese by length heterogeneity-PCR. J. Appl. Microbiol. 96:481–490.
Lodge-Ivey, S.L., M.S. Rappe, W.H. Johnston, R.E. Bohlken, and A.M. Craig. 2005. Molecular analysis of a consortium of ruminal microbes that detoxify pyrrolizidine alkaloids. Can. J. Microbiol. 51:455–465.
Lord, N.S., C.W. Kaplan, P. Shank, C.L. Kitts, and S.L. Elrod. 2002. Assessment of fungal diversity using terminal restriction fragment (TRF) pattern analysis: Comparison of 18S and ITS ribosomal regions. FEMS Microbiol. Ecol. 42:327–337.
Matsumoto, M., M. Sakamoto, H. Hayashi, and Y. Benno. 2005. Novel phylogenetic assignment database for terminal-restriction fragment length polymorphism analysis of human colonic microbiota. J. Microbiol. 61:305–319.
Mills, D.K., K. Fitzgerald, C.D. Litchfield, and P.M. Gillevet. 2003. A comparison of DNA profiling techniques for monitoring nutrient impact on microbial community composition during bioremediation of petroleum contaminated soils. J. Microbiol. 54:57–74.
Mills, D., J. Entry, J. Voss, and K. Mathee. 2006. An assessment of the hypervariable domains of the 16S rRNA genes for their value in determining microbial community diversity: The paradox of traditional ecological indices. FEMS Microbiol. Ecol. 57:496–503.
Moeseneder, M.M., J.M. Arrieta, G. Muyzer, C. Winter, and G.J. Herndl. 1999. Optimization of terminal-restriction fragment length polymorphism analysis for complex marine bacterioplankton communities and comparison with denaturing gradient gel electrophoresis. Appl. Environ. Microbiol. 65:3518–3525.
Muyzer, G., E.C. DeWaal, and A.G. Uitterlinden. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. 59:695–700.
Norton, J.M., J.J. Alzerreca, Y. Suwa, and M.G. Klotz. 2002. Diversity of ammonia monooxygenase operon in autotrophic ammonia-oxidizing bacteria. Arch. Microbiol. 177:139–149.
Pace, N.R. 1997. A molecular view of microbial diversity and the biosphere. Science 276:734–740.
Ranjard, L., F. Poly, J.-C. Lata, C. Mougel, J. Thioulouse, and S. Nazaret. 2001. Characterization of bacterial and fungal soil communities by automated ribosomal intergenic spacer analysis fingerprints: Biological and methodological variability. Appl. Environ. Microbiol. 67:4479–4487.
Ritchie, N.J., M.E. Schutter, R.P. Dick, and D.D. Myrold. 2000. Use of length heterogeneity PCR and fatty acid methyl ester profiles to characterize microbial communities in soil. Appl. Environ. Microbiol. 66:1668–1675.
Sekar, R., D.K. Mills, E.R. Remily, J.D. Voss, and L.L. Richardson. 2006. Microbial communities in the surface mucopolysaccharide layer and the black band microbial mat of back band-diseased *Siderastrea siderea*. Appl. Environ. Microbiol. 72:5963–5973.
Simpson, J.M., J.W.S. Domingo, and D.J. Reasoner. 2002. Microbial source tracking: State of the science. Environ Sci. Technol. 36:5279–5288.
Stephen, J.R., Y.-J. Chang, Y.D. Gan, A. Peacock, S.M. Pfifner, M.J. Barcelona, D.C. White, and S.J. Macnaughton. 1999. Microbial characterization of a JP-4 fuel-contaminated site using a lipid biomarker/polymerase chain reaction–denaturing gradient gel electrophoresis (PCR–DGGE)-based approach. Environ. Microbiol. 1:231–241.
Suzuki, M.T., and S.J. Giovannoni. 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. 62:625–630.
Suzuki, M., M.S. Rappe, and S.J. Giovannoni. 1998. Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. Appl. Environ. Microbiol. 64:4522–4529.
Swift, M.J., A.-M.N. Izac, and M. van Noordwijk. 2004. Biodiversity and ecosystem services in agricultural landscapes: Are we asking the right questions? Agric. Ecosyst. Environ. 104:113–134.
Tiirola, M.A., J.E. Suvilampi, M.S. Kulomaa, and J.A. Rintala. 2003. Microbial diversity in a thermophilic aerobic biofilm process: Analysis by length heterogeneity PCR (LH-PCR). Water Res. 37:2259–2269.
Torsvik, V., R. Sorheim, and J. Goksoyr. 1996. Total bacterial diversity in soil and sediment communities: A review. J. Ind. Microbiol. 17:170–178.
Venter, J.C., K. Remington, J. Heidelberg, A.L. Halpern, D. Rusch, J.A. Eisen et al. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science 304:66–74.
Vestal, J.R., and D.C. White. 1989. Lipid analysis in microbial ecology. BioScience 39:535–541.
Watnick, P., and R. Kolter. 2000. Biofilm, city of microbes. J. Bacteriol. 182:2675–2679.
Watve, M.G., and R.M. Gangel. 1996. Problems in measuring bacterial diversity and a possible solution. Appl. Environ. Microbiol. 62:4299–4301.
Wintzingerode, F.V., U.B. Gobel, and E. Stackenbrandt. 1997. Determination of microbial diversity in environmental samples: Pitfalls of PCR-base rRNA analysis. FEMS Microbiol. Rev. 21:213–229.
Wolters, V., W.L. Silver, D.E. Bignell, D.C. Coleman, P. Lavelle, W.H. van der Putten et al. 2000. Effects of global changes on above- and belowground biodiversity in terrestrial ecosystems: Implications for ecosystem functioning. BioScience 50:1089–1098.
Xing, J., C. Criddle, and R. Hickey. 1997. Long-term adaptive shifts in anaerobic community structure in response to a sustained cyclic substrate perturbation. FEMS Microbiol. Ecol. 33:50–58.
Yang, C., D. Mills, K. Mathee, Y. Wang, K. Jayachandran, M. Sikaroodi, P. Gillevet, J. Entry, and G. Narasimhan. 2006. An ecoinformatics tool for microbial community studies: Supervised classification of amplicon length heterogeneity (ALH) profiles of 16S rRNA. J. Microbiol. 65:49–62.